Introduction

Many claim that the sizes of major interstate conflicts follow the power-law, a model that characterizes phenomena with thick-tailed distributions. Such data are difficult to study with conventional statistics, however a number of innovative methods for making inferences and testing hypotheses about trends in the sizes of major wars using the power-law model have been introduced in the past decade (Braumoeller 2019; Cederman, Warren, and Sornette 2011; Cirillo and Taleb 2016; Clauset 2017, 2018; Cunen, Hjort, and Nygård 2020). Despite these innovations, studying data with the power-law also usually requires data truncation to ensure the model optimally fits the extreme tail of a thick-tailed distribution (Clauset, Shalizi, and Newman 2009). This can result in effective data loss in excess of 50%, leaving variation in a majority of conflicts unexplained and hurting statistical precision for the remaining data.[1] Are these limitations necessary? Are other, more conventional, modeling choices really unjustifiable?

Answers to these questions are worth seeking because studies that use the power-law to study the sizes of wars do so to answer questions of serious normative importance. How likely is a given war to result in the deaths of a few thousand or more than one million? Is there such a thing as a *long peace* that began in the mid-20th century? Are there correlates of war size, or is war size entirely random? Given recent escalations in many conflicts across the globe, these questions are more timely than ever. According to the Uppsala Conflict Data Program, the years 2022 and 2023 have witnessed more global conflicts than any other year since the end of the Cold War.[2] Yet, at a time when data-driven guidance is needed, current research yields conflicting conclusions. Where some fail to find evidence of statistically detectable changes in war size over time (Cirillo and Taleb 2016; Clauset 2017, 2018; Braumoeller 2019), others do (Cunen, Hjort, and Nygård 2020). And where some argue that war size is mostly random (Braumoeller 2019), others claim that factors like democracy can explain why some wars are deadlier than others (Cunen, Hjort, and Nygård 2020).

The approaches taken by prior studies are not uniform, but what remains the same is a commitment to the idea that truly large wars are power-law distributed. However, as scholars across a wide range

[1] There are some solutions around this problem, but these are rarely applied (at least within the peace science literature).
[2] Accessed Nov. 20, 2023: https://ucdp.uu.se/

of fields are quickly discovering, just because the power-law model can be fit to data does not imply that it should. Recent research shows that many events such as solar flares and city population sizes that were long-thought to follow the power-law are better treated as log-normal (Fazio and Modica 2015; Verbeeck et al. 2019). Perhaps it is time to question the usefulness of the power-law for studying war size as well.

Best-practices for studying data suspected of following the power-law exist (Clauset, Shalizi, and Newman 2009), but these procedures are rarely applied in studies of war size. This is remedied in this study, and alongside using these methods, the appropriateness of two alternative models is also investigated: the log-normal model and the inverse Burr model (a power-law friendly but more flexible distributional form). Unlike the classic power-law, each of these approaches can be used to fit an entire dataset (not just the most extreme events) and is amenable to the inclusion of covariates. The findings of this study show that among these three alternatives, scholars have more freedom of choice than previously thought.

It bears noting that the appropriateness of the power-law for explaining conflict has been broached before (González-Val 2016; Zwetsloot 2018). This study goes beyond these studies in two important ways. First, it considers the importance of how war size is measured. Using the Correlates of War interstate conflict series, the analysis undertaken here shows that the choice to normalize war size by population (or not) alters conclusions about the best fitting model. Three ways of measuring of war size are considered: (1) *severity* of total battle deaths, (2) *prevalence* of deaths per global population, and (3) *intensity* of deaths per the populations of the countries fighting a war.[3]

Across measures, the classic power-law cannot be formally rejected as a good fit for the data. However, to properly fit the extreme tail of the distribution 50-80% data truncation is required, regardless of measure. The performance of the log-normal and inverse Burr models, both of which can be used without data truncation, varies depending on the measure in question. Neither the log-normal nor inverse Burr are a supportable alternative when measuring conflict by absolute *severity*. However, the inverse Burr cannot be rejected as a good fit when measuring conflict deaths by global *prevalence*,

[3]This is the typology proposed by Braumoeller (2019).

and both the inverse Burr and log-normal cannot be rejected when measuring deaths by per capita *intensity*. In both cases, while the power-law only fits less than half the data, the inverse Burr and log-normal can explain *all* the data.

In addition to showing that conflict scholars have freedom of choice in studying conflict sizes, the second contribution of this study is that it shows that this freedom of choice is not free of consequence. Further analysis demonstrates that the choice of model and measure determines whether a *long peace* starting in the mid-20th century is statistically detectable, whether correlates of war size can be easily parameterized (and thus controlled for), and whether said correlates have statistically significant relationships with war size. With respect to implications for inferences, a *long peace* is not statistically detectable using the classic power-law or inverse Burr models, regardless of how war size is measured. Conversely, a *long peace* is statistically detectable for war death *intensity* when using the log-normal model. With respect to modeling correlates of war size via regression analysis, both the log-normal and inverse Burr models can support the inclusion of covariates like population, military size, and democracy. Both approaches yield similar coefficients but provide slightly different inferences about the statistical significance of predictors of war size. Generally, a larger number of military personnel predicts a higher number of battle deaths, as does a larger global population. Meanwhile, a larger population of the countries fighting a war, as well as higher quality of democracy, predicts fewer battle deaths. However, these findings are not consistent across measures of conflict size or depending on whether the model was inverse Burr or log-normal.

Taken as a whole, this study raises important issues and complications for peace scholars to consider. First, the power-law model should not be taken for granted when studying war size. Depending on whether size is measured in absolute terms or normalized by population, the inverse Burr or log-normal will be just as appropriate, explain more of the data, provide better statistical precision, and permit parameterization of covariates. Second, since the choice of measure will determine model selection, by extension it will determine results. This sensitivity to measurement and model selection demands further investigation, and some advice for how conflict scholars should proceed is provided in the concluding section.

The paper proceeds as follows. First, in the next section some of the relevant properties of the classic power-law model and alternatives are discussed. Then, the recommended "recipe" for estimating, validating, and comparing these models is summarized. Next, the data used (Correlates of War interstate conflict series) for model fitting are discussed. Finally, the results are presented, followed by a discussion of implications and recommendations.

The Power-law and Alternatives

One of the most pressing questions in the quantitative study of war is how to explain variation in the sizes of international conflicts. International wars are said to follow Richardson's Law, which holds that most wars kill only a few combatants while a few are likely to be exceptionally deadly. The discovery of this regularity is owed, in part, to early contributions to the quantitative study of conflict by Lewis F. Richardson (1948, 1960) who compiled original data on the size and duration of historical international conflicts.

Evidence of Richardson's Law persists in more up-to-date and now well-established datasets such as the Correlates of War (CoW) inter-state conflict series, which documents the battle deaths from 95 interstate wars fought between 1816 and 2007 (Sarkees and Wayman 2010). Figure 1 shows the distribution of total battle deaths from the 95 wars in the dataset. For ease of interpretation, battle deaths are shown on the log-10 scale. It is plain to see that the distribution abides by Richardson's Law. The bottom 80% of wars in terms of deadliness account for only 1.01% of total battle deaths in the data. Conversely, the top 20% of wars are responsible for 98.99% of battle related deaths in interstate conflicts. That is a remarkable disparity.

To model this unique distribution of battle deaths from interstate conflicts, researchers have typically turned to the power-law. Power-law generated data display characteristically thick extreme tails such as those seen in the CoW conflict series. The power-law model characterizes the inverse cumulative distribution function (CDF), or the probability of an event of size X greater than x as

$$\Pr(X > x) \propto x^{-\alpha} \quad \text{for all large } x \tag{1}$$

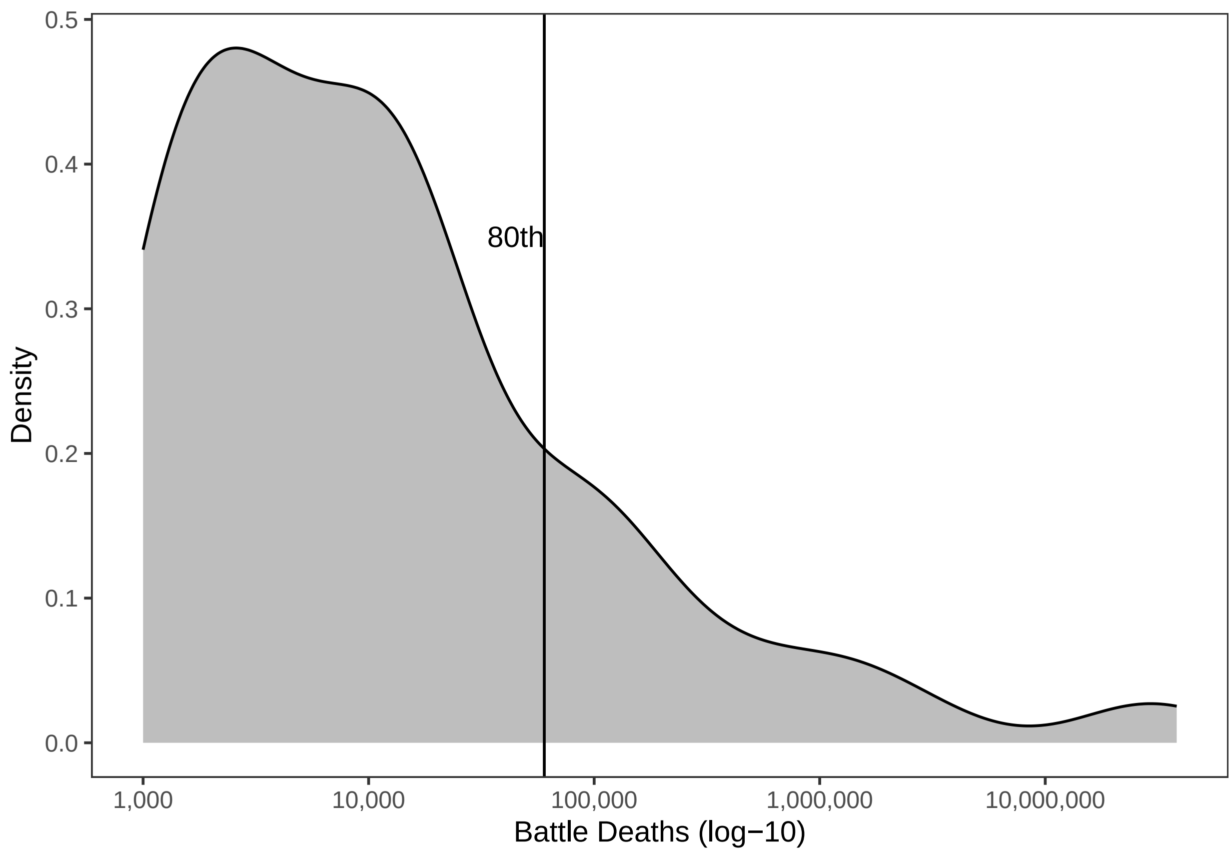

Figure 1: Density plot of total battle deaths from CoW battle series, 1816-2007. The x-axis is on the log-10 scale. The 80th percentile is denoted with a vertical line.

where $\alpha > 0$. That is, the probability of an event $X > x$ is inversely proportional to the size of the event raised to the power α. As $\alpha \to 0$, the tail of the distribution becomes thicker, meaning the likelihood of even extremely large events is quite high.

The power-law model has many unique properties, including linearity between the inverse CDF and observed event size on the log-log scale. That is:

$$\log[\Pr(X > x)] \propto -\alpha \log(x). \tag{2}$$

This is illustrated in Figure 2, which compares the theoretical inverse CDF of a hypothetical variable x on an unadjusted scale versus a log-log scale. This characteristic of the classic power-law model often is why early efforts to estimate α with empirical data relied on OLS, an approach that has since been shown to be unreliable in some circumstances. The current recommended practice is to use the maximum likelihood estimator (MLE) summarized by Clauset, Shalizi, and Newman (2009).

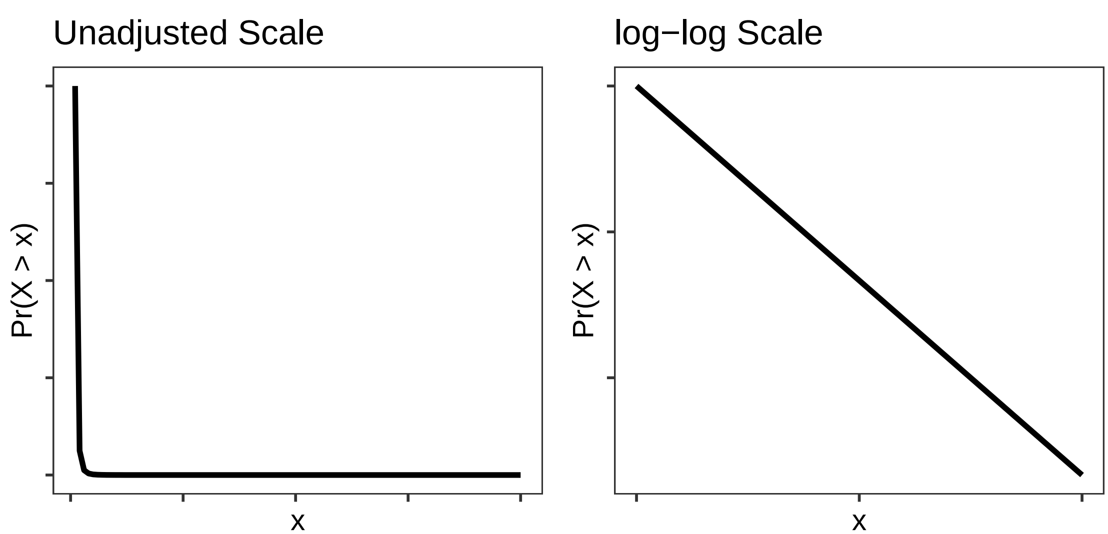

Figure 2: The inverse CDF of power-law data in unadjusted scale versus log-log scale.

In practice, linearity in log-log space is rarely so consistent across the entire set of observed data. For example, using the CoW conflict series, but this time with battle deaths adjusted to the population size of the countries fighting a war, the relationship between the empirical CDF and x in log-log space displays clear quasi-concavity (Figure 3). This is true for many other phenomena where sometimes we only observe this characteristic linearity in the extreme tail of the distribution, giving rise to the necessity of identifying $x_{\min}$ such that all $x \geq x_{\min}$ are power-law distributed.

This step of identifying $x_{\min}$ is the state-of-the-art for fitting the power-law to data (Clauset, Shalizi, and Newman 2009). The consequences can sometimes be minimal, but in other cases this approach can lead to substantial data loss. However, this can be justified if we really believe the data are power-law distributed in the extreme tail, and if such events are the primary focus of study.

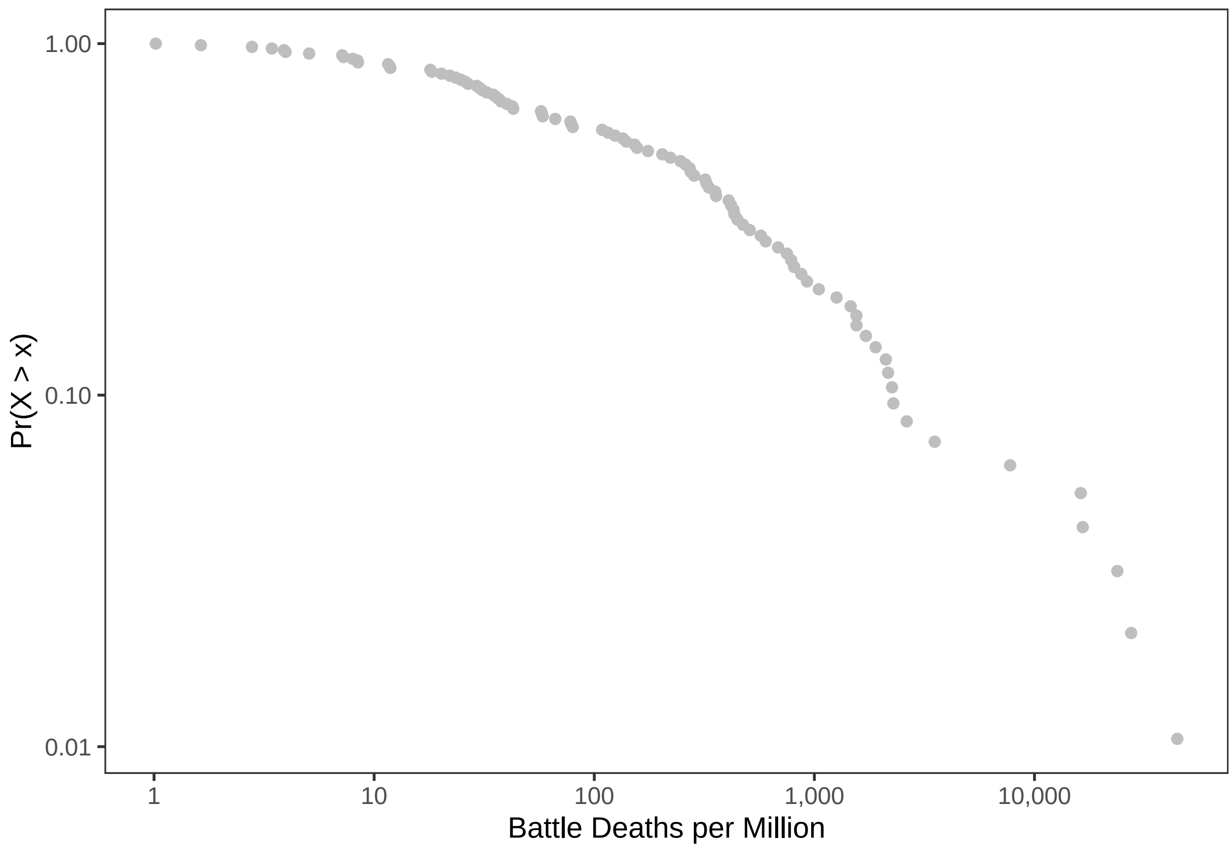

Figure 3: The inverse ECDF of battle deaths per the populations of the countries fighting a war per 1 million shown in the log-10 scale.

Studying the extreme tails of phenomena with the power-law model comes with some provocative implications. Most notable among these is the possibility of identifying scale-free phenomena. If $\alpha \leqslant 3$, the data lack finite variance, and if $\alpha \leqslant 2$ the data lack a finite mean. In such extreme cases, the phenomenon under study can be subject to *black swan* behavior—events that are exceptionally extreme and inexplicable—with the expected magnitude of such events statistically indistinguishable from infinity (Taleb 2010). This has made the power-law especially relevant to conflict scholars interested in studying the deadly potential of international war, and recent research on the CoW battle series in fact finds that $\alpha < 2$ (Braumoeller 2019). Such a finding supports the conclusion that interstate conflicts (worryingly) have black swan tendencies.

Of course, the classic power-law model is not the only one that can capture data with a skewed tail. There are alternatives that have more favorable, though less provocative, properties for statistical analysis. In an excellent study, Cunen, Hjort, and Nygård (2020) recently used a more general distributional form known as the inverse Burr to study the CoW battle series summarized above. The inverse Burr distribution specifies the probability of an event greater than size x as:

$$\Pr(X > x) = 1 - \left[\frac{(x/\mu)^\theta}{1 + (x/\mu)^\theta}\right]^\alpha \tag{3}$$

where the parameters μ, θ, and α are strictly greater than zero. The parameter μ is a scaling parameter that captures the central tendency of x, while θ and α are shape parameters. θ in particular functions much the same way that α does in characterizing the extreme tails of power-law distributed data. This is because as x increases we have:

$$\Pr(X > x) \approx \alpha(\mu/x)^\theta. \tag{4}$$

According to Cunen, Hjort, and Nygård (2020), the strength of the inverse Burr relative to the classic power-law is its ability to model the entire conflict series, not just the extreme tail, when the relationship between the inverse empirical CDF and the data is quasi-concave in log-log space. This ability is on display in Figure 4, which shows the relationship between the inverse CDF and some hypothetical observed data assuming an inverse Burr distribution in the log-log scale. Note the characteristic downward curve.

With this increased flexibility comes greater statistical power because the model can be fit efficiently with all the data, not just the most extreme events. This is the justification that Cunen, Hjort, and

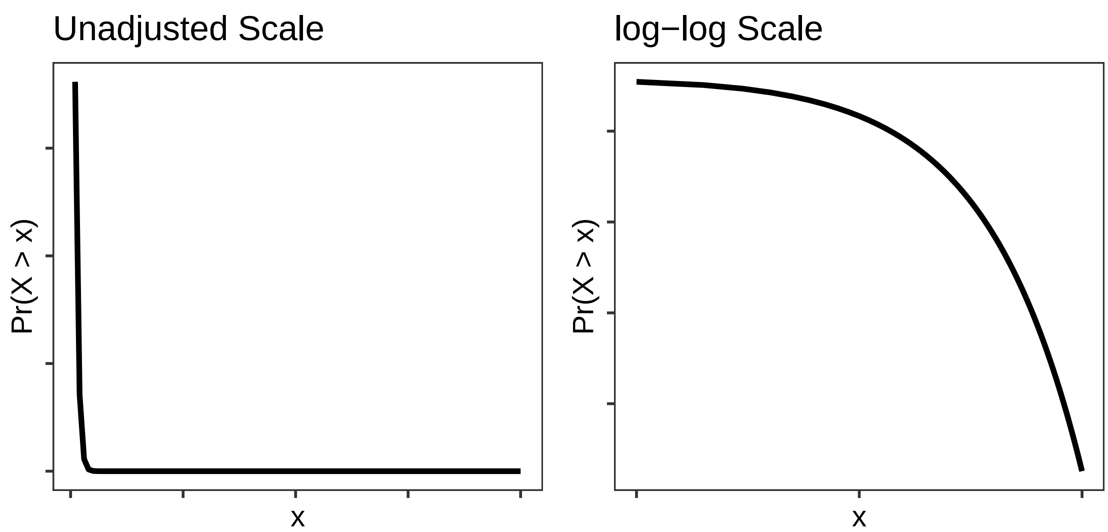

Figure 4: The inverse CDF of inverse Burr data in unadjusted scale versus log-log scale.

Nygård (2020) make for preferring the inverse Burr over the classic power-law in their analysis of the CoW battle data; however, the authors do not report the results of a formal goodness of fit test to formally validate their modeling choice.

Importantly, the inverse Burr model is not the only distributional form that can more flexibly model both smaller and larger events simultaneously, nor is it necessarily the most useful. The log-normal distribution, as the name suggests, characterizes data that is normally distributed in the log-scale. It has the inverse CDF:

$$\Pr(X > x) = 1 - \Phi([\log(x) - \mu]/\sigma) \tag{5}$$

where μ and $\sigma > 0$ are the log-mean and log-standard deviation, respectively, and $\Phi(\cdot)$ is the normal CDF.

Like the inverse Burr, the log-normal model can capture a curved relationship between the inverse CDF and the data in log-log space, as shown in Figure 5. Note that the fit is not identical to that of the inverse Burr. It is slightly more severe, which leads the log-normal model to give a slightly higher probability to smaller events and a lower probability to larger events.

The log-normal and inverse Burr have three advantages over the classic power-law, assuming they provide a good fit for the data. First, as already noted, each can be fit to the entire conflict series.

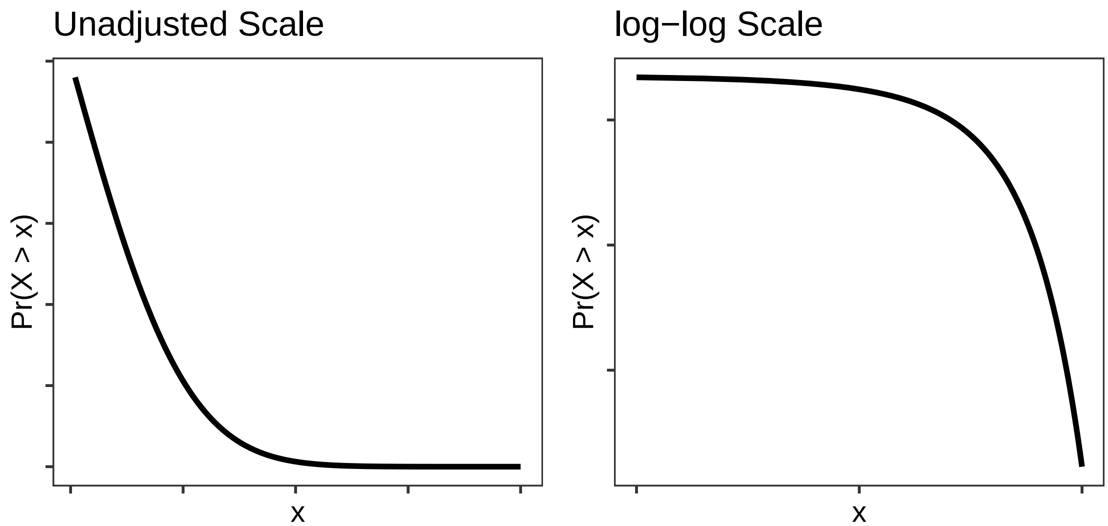

Figure 5: The inverse CDF of log-normal data in unadjusted scale versus log-log scale.

Second, for all $x > 0$ each has consistently identifiable parameters with finite variance. This means each provides stable estimates of the probabilities of various war sizes (e.g., no expected values in the extreme tail that are statistically indistinguishable from infinity). Third, each can incorporate covariates in model estimation—a strength illustrated by Cunen, Hjort, and Nygård (2020) using the inverse Burr model.

Importantly, the log-normal has the greatest advantage with respect to the second and third points raised above. Nearly all statistical software comes pre-packaged with tools for doing statistical inference and regression analysis using log-normal data. Similar tools for working with inverse Burr models may not always be available. It is, however, possible to program these tools by hand, as authors like Cunen, Hjort, and Nygård (2020) do in their study. This is also done for this study, and readers interested in the programming details can find them in the Code Appendix.[4]

[4]The R code creates a function called `inbur_reg()` that accepts a formula object and data just like conventional modeling functions, such as `lm()` and `glm()` in base R.

Methods for Model Fitting and Selection

Clauset, Shalizi, and Newman (2009) lay out a simple "recipe" for analyzing data with the power-law model. Their set of best-practices is adopted here to compare the power-law to the inverse Burr and log-normal models discussed in the previous section.

The first step in the recipe is simply to fit the power-law, as well as the other models, to the data. The second step is to perform goodness of fit (GOF) tests. Clauset, Shalizi, and Newman (2009) detail a bootstrapping/simulation approach that involves first estimating a distance metric such as the KS (Kolmogorov-Smirnov) statistic for the fitted model. Next, using the fitted model, simulate new synthetic datasets. The model is then fit to the synthetic data and a KS statistic is calculated using the synthetic data and the model parameters fit to that specific synthetic dataset. A p-value from this test is calculated as the share of times the KS distances measured using the simulated data are larger than the empirical distance. If the p-value is small then the model in question is not a plausible fit for the data. Clauset, Shalizi, and Newman (2009) recommend $p < 0.1$ as the cutoff.

The third and final step involves direct comparisons of the models being considered. This may not be necessary if one or more models are rejected in step 2. However, in some cases separate GOF tests will not be definitive. It is possible that two or more models provide a good fit and none can be rejected. When this is the case, Clauset, Shalizi, and Newman (2009) suggest a likelihood ratio (LR) test to formally judge between competing models. In particular, they recommend Vuong's test, which is an LR-based test for model selection using Kullback-Leibler criteria (see Vuong 1989). This is a non-nested test, which is important to use since the models in question are not nested versions of one another. The sign of the test will indicate which model is the better fit (the better model will have a higher likelihood), and the p-value will indicate whether we should reject the null that the models perform the same.

A caveat with the LR test is that it requires calculating observation specific likelihoods. This means that if one model is fit with an entire dataset (say the log-normal) but another is only valid for all $x \geq x_{\text{min}}$ (the power-law), only the likelihoods for all $x \geq x_{\text{min}}$ can be used to perform the LR test. This means that if one model fits $x < x_{\text{min}}$ as well, this will not be factored into the test.

In sum, the recipe for analyzing the data will consist of three steps:

1. Fit each of the models (classic power-law, inverse Burr, and log-normal) to the data.
2. Perform GOF tests to see if any of the models can be formally rejected.
3. Perform LR tests to formally assess whether one model is a better fit for the data than another.

The next section outlines the data and measures that will be used for model estimation, validation, and comparison.

Data and Measures

The data used for the analysis comes from the CoW interstate conflict dataset, which documents battle deaths per country across 95 interstate wars fought between 1816 and 2007 (Sarkees and Wayman 2010; D. J. Singer 1987). The data were accessed using the `{peacesciencer}` R package (Miller 2022). For each conflict in the dataset, the total number of battle deaths across countries fighting a war were tallied, yielding a conflict series of 95 observations.

Other datasets have been assembled of historical wars, some going as far back as the 1400s (Cederman, Warren, and Sornette 2011) and others to 1 A.D.(Cirillo and Taleb 2016). The CoW conflict series is chosen for two reasons. The first is its widespread use, which makes comparisons with approaches used in other studies easier (Braumoeller 2019; Cederman 2003; Clauset 2017, 2018; Cunen, Hjort, and Nygård 2020). The second justification is that the CoW data are generally considered of good quality. That does not mean they are regarded as perfect. As previously mentioned, the data contain some irregularities, and for some conflicts death counts are disputed (Reiter, Stam, and Horowitz 2016). The data further have limited coverage. While the conflict series runs from 1816 to 2007 (nearly two centuries worth of wars), the history of human conflict did not start in 1816, nor did it end in 2007. These limitations aside, the benefits noted above make the CoW series a best choice among imperfect alternatives.

With a dataset chosen, the next issue to settle is how to measure war size. Should total battle deaths be studied, or should deaths be normalized by population? Braumoeller (2019) provides a helpful typology that summarizes the alternative approaches. He specifically denotes three: (1) *severity* of war deaths, (2) *prevalence* of war deaths, and (3) *intensity* of war deaths. The first, *severity*, measures war size in absolute magnitude. The second and third measures of war size are in relative terms. *Prevalence* normalizes war deaths by global population at the time of a conflict, while *intensity* does so by the populations of the countries fighting a war.

There is no right measure. Which is best depends on what question we want to answer. *Severity* is useful if we care about how many people in total are likely to die in a war, while *prevalence* and *intensity* are useful if we care about the relative risk of dying in war. When considering relative risk, some scholars prefer to use global population as the denominator (Pinker 2011) while others prefer to limit the denominator to the populations of the countries involved in a war (Braumoeller 2019). This choice follows from different goals. Deaths per global population (*prevalence*) is akin to a measure of all-cause mortality risk from war, meaning it treats war like a public health issue. Deaths per the populations of the countries fighting a war (*intensity*) treats war like a political activity or behavior that has unique consequences for the countries involved.

Rather than belabor the merits of one approach over another, the choice is made here to use all three, since each quantity can be of interest depending on the research question being asked. It is also possible that different measures will yield different results about model fit. If this is the case, conflict scholars would obviously want to take note.

The dataset to be analyzed, then, has for each given war i in $i = 1, ..., 95$ a measure of *severity*, a measure of *prevalence*, and a measure of *intensity*. The normalized measures are scaled to deaths per million to make the results more intuitive. The data for population come from version 6.0 of the CoW National Military Capabilities dataset and were also accessed using the `{peacesciencer}` package (D. J. Singer 1987; J. D. Singer, Bremer, and Stuckey 1972; Miller 2022).

Analysis

Having established the conflict series and measures of war size to be used, the analysis proceeds according to the recipe outlined previously in the paper. The first subsection below shows the results from model fitting. The next shows the results from GOF tests. The third shows the results from LR tests comparing model fits.

Model Fitting

Figure 6 provides a visual representation of the model fits for each of the measures of war size (severity, prevalence, and intensity). Each panel is a scatter plot that shows the relationship between the observed sizes of wars in the conflict series (x-axis) and the in inverse ECDF of war size (y-axis). Values are shown on the log-log scale. The black line denotes the power-law fit for all $x \geq x_{\text{min}}$, the blue curve denotes the inverse Burr fit, and the red curve denotes the log-normal fit.

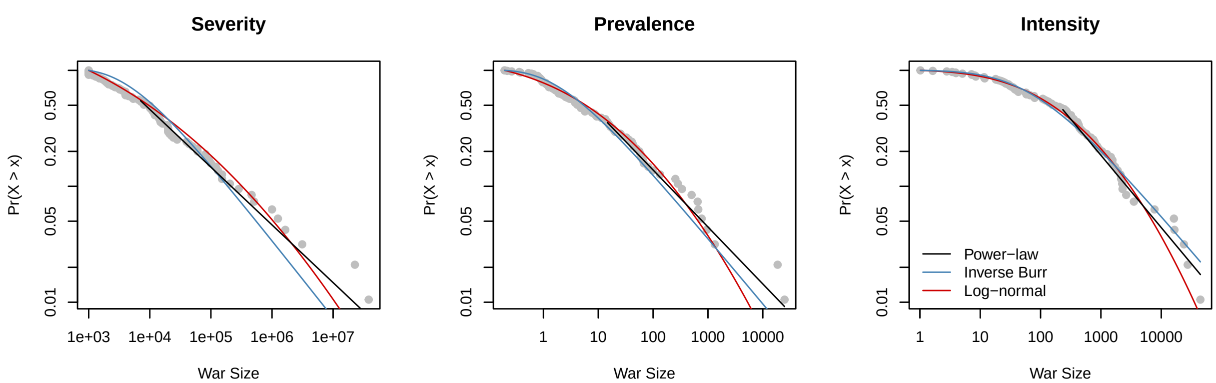

Figure 6: Visualization of model fits for the data. The empirical inverse CDF is shown over the data in log-log space across panels. The first panel shows results for total battle deaths. The second shows results for battle deaths per global population in millions. The third shows results for battle deaths per population of the countries at war in millions.

For *severity* of war deaths, the best-fitting power-law slope applies to all wars that resulted in at least 7,061 battle deaths (54.74% of observations) with a slope coefficient of 1.5. This is consistent with war *severity* being scale-free in the extreme tail of the distribution, which echos previous findings using the CoW conflict series.

Both the inverse Burr and log-normal models were fit to the entire data sample. As shown in Figure 6, the fits in the extreme tail deviate from the power-law fit in interesting ways. The inverse Burr under-predicts the probability of some of the most severe conflicts, even compared to the log-normal model. Meanwhile, it over predicts the likelihood of smaller conflicts relative to the log-normal. The fitted parameters for the inverse Burr are 611255.323, 0.669, and 0 for the two shape parameters and the scaling parameter respectively. The fitted parameters for the log-normal are 5.777 and 3.91 for the log-mean and log-standard deviation.

For *prevalence* of war deaths, the best-fitting power-law slope applies to all wars that resulted in at least 14 battle deaths (36.84% of observations) with a slope coefficient of 1.49. Again, this is consistent with war size being scale-free in the extreme tail of the distribution.

As shown in the middle panel of Figure 6, the fits for the inverse Burr and log-normal in the extreme tail deviate less from the power-law fit compared to *severity*. However, the inverse Burr still under-predicts the probability of some of the most severe conflicts relative to the power-law, as does the log-normal. The fitted parameters for the inverse Burr are 752061.415, 0.567, and 0 for the two shape parameters and the scaling parameter respectively. The fitted parameters for the log-normal are 0.957 and 3.157 for the log-mean and log-standard deviation.

Finally, for *intensity* of war deaths, the best-fitting power-law slope applies to all wars that resulted in at least 222 battle deaths (84.21% of observations) with a slope coefficient of 1.62. Yet again, this is consistent with war size being scale-free in the extreme tail of the distribution.

As shown in the left panel of Figure 6, the fits for the inverse Burr and log-normal in the extreme tail deviate very little from the power-law fit. The inverse Burr slightly over-predicts the probability of some of the most severe conflicts relative to the power-law and the log-normal. The fitted parameters for the inverse Burr are 2.48, 0.614, and 21.102 for the two shape parameters and the scaling parameter respectively. The fitted parameters for the log-normal are 4.985 and 2.353 for the log-mean and log-standard deviation.

Table 1: GOF Tests for Severity

Model	GOF	p-value
Power-law	0.063	0.874
Inverse Burr	0.135	0.094
Log-normal	0.614	0.000

Based on 2,000 bootstraps.

Goodness of Fit

Visual inspection of the models creates the impression that, while each fits much of the data relatively well, the quality of fit improves for the inverse Burr and log-normal when war size is normalized by population. This especially looks to be the case when war size is measured as *intensity*. This section discusses the results from formal GOF tests that provide a more precise estimate for how well these models fit each of the measures.

Table 1 shows estimates from the simulation-based GOF test described previously in the paper for each of the models fit using conflict severity. Recall that these tests infer goodness of fit by comparing how well a model fits the data assuming the model were the true data-generating process. For each of the models a KS distance statistic was calculated. Then multiple simulated datasets were generated using the estimated model parameters. The same model was then fit to each of the simulated datasets and the KS distance for that dataset calculated. A p-value was then computed by calculating the fraction of the simulations where the distance metric was greater than the one calculated with the original model estimates and data.

The results show that when conflict is measured in terms of absolute severity, only the classic power-law model cannot be formally rejected: $D = 0.06$ (p-value = 0.87). Conversely, the inverse Burr and log-normal can be rejected. Both have p-values < 0.1 (though the inverse Burr just narrowly crosses this threshold), which is the cutoff recommended by Clauset, Shalizi, and Newman (2009). That we can reject the null that the data follow the inverse Burr is especially notable since this was the preferred method of Cunen, Hjort, and Nygård (2020) in studying the same CoW data used here. The authors in particular measured war in terms of absolute severity. The results presented here do

Table 2: GOF Tests for Prevalence

Model	GOF	p-value
Power-law	0.077	0.847
Inverse Burr	0.088	0.482
Log-normal	0.259	0.004

Based on 2,000 bootstraps.

Table 3: GOF Tests for Intensity

Model	GOF	p-value
Power-law	0.072	0.805
Inverse Burr	0.073	0.137
Log-normal	0.057	0.640

Based on 2,000 bootstraps.

not support their approach. When the goal is to study war severity, the classic power-law is the only approach that cannot be rejected.

The results differ when we consider the global prevalence of war deaths. Table 2 reports the GOF test results for all the models when fit to prevalence. Both the classic power-law and inverse Burr models cannot be rejected while the log-normal can. The p-values associated with each of the tests are 0.85, 0.48, and 0 respectively. Importantly, while the power-law cannot be rejected in the extreme tail of the data, the inverse Burr cannot be rejected as the data-generating process for all the data.

The results differ yet again when considering war intensity. Table 3 reports GOF test results for each of the models fit using the intensity of war deaths per the populations of the countries fighting a war. Here, none of the models can be rejected. The p-values associated with each of the tests are 0.81, 0.14, and 0.64 respectively. Again, it is important to note that while the classic power-law cannot be rejected in the extreme tail, the log-normal and inverse Burr cannot be rejected while also explaining all of the data.

Model Comparisons

The results up to now show that the hypothesis that the data are power-law distributed in the extreme tail of the distribution cannot be rejected, regardless of whether war size is measured in terms of absolute severity or in relative terms (prevalence or intensity). However, alternatives also cannot be definitively rejected when considering battle death prevalence and intensity. In the former case, the inverse Burr cannot be rejected while in the latter case, the log-normal and the inverse Burr cannot be rejected. Clearly, since different alternatives are plausible fits for the data, additional tests are required to formally adjudicate which is best.

This section shows results from Vuong's (1989) likelihood ratio (LR) test, which is a nonnested model comparison test. As noted in the methodological discussion in a previous section, a limitation of this approach is that it does not factor in the data truncation imposed by using the power-law model as compared to the alternatives when comparing model fit. However, at minimum, it provides a way to quantify how well the alternative models fit the data in the extreme tail of the the distribution.

Up first, Table 4 shows the results from Vuong's test applied to pairwise comparisons of the three competing models for conflict severity. Recall that only the power-law could not be rejected in the previous section, so from a data-driven perspective there already are grounds for selecting the power-law for studying war severity in lieu of the log-normal or inverse Burr distributions. Even so, the model comparisons are worth mentioning, because according to Vuong's test the power-law is *not* statistically better than either the log-normal or inverse Burr in the extreme tail of the distribution; albeit, the sign of the test statistic is positive, which means the likelihood for the power-law is greater than the others. Between the inverse Burr and the log-normal, the null of Vuong's test can be rejected. The negative sign on the test indicates that the inverse Burr performs worse than the log-normal; however, again, both of these models were nonetheless rejected in the previous set of GOF tests.

Table 5 shows results from Vuong's test when modeling battle death prevalence. In this case, none of the models can be rejected in favor of another. Recall again than in the previous GOF tests only the log-normal model could be rejected, meaning regardless of the results here, we do have a basis for narrowing down our options to either the power-law or inverse Burr. The signs on the test statistics

Table 4: Vuong's test for best fitting model for battle death severity.

Models	Estimate	p-value
Power-law vs. Inverse Burr	0.854	0.393
Power-law vs. Log-normal	0.864	0.388
Inverse Burr vs. Log-normal	-2.384	0.017

Only non-truncated data points used for comparisons with the power-law. Full data used for inverse Burr vs. log-normal test.

Table 5: Vuong's test for best fitting model for battle death prevalence.

Models	Estimate	p-value
Power-law vs. Inverse Burr	0.064	0.949
Power-law vs. Log-normal	0.343	0.732
Inverse Burr vs. Log-normal	0.410	0.681

Only non-truncated data points used for comparisons with the power-law. Full data used for inverse Burr vs. log-normal test.

indicate that the power-law does perform slightly better relative to the alternatives, but this difference is not statistically different from zero.

Finally, Table 6 shows results from Vuong's test applied to models fit to battle death intensity. Recall from the previous section that only in the case of war intensity was the null not rejected for each of the competing models. The failure to definitively reject at least one of the alternatives in the GOF tests makes the results from Vuong's test all the more vital in model selection. The estimates here seem to favor the log-normal model. In comparisons with the power-law and inverse Burr, the log-normal's likelihood in the extreme tail of the distribution is higher. Unfortunately, as was the case in modeling battle death prevalence, this better performance is not statistically significant.

Implications

The foregoing analysis shows that there is no definitive best model of war size among the three considered. While only the power-law model cannot be rejected in the case of absolute battle death

Table 6: Vuong's test for best fitting model for battle deaths intensity.

Models	Estimate	p-value
Power-law vs. Inverse Burr	0.328	0.743
Power-law vs. Log-normal	-0.153	0.878
Inverse Burr vs. Log-normal	-1.373	0.170

Only non-truncated data points used for comparisons with the power-law. Full data used for inverse Burr vs. log-normal test.

severity, it and the inverse Burr also cannot be rejected for battle death prevalence, and all three models (the power-law, inverse Burr, and log-normal) cannot be rejected in the case of war death intensity. Additional model comparison tests fail to provide a data-driven answer for which of the surviving alternatives should be preferred. While the signs of the tests favor the power-law in the case of severity and prevalence and the log-normal in the case of intensity, the test statistic is not statistically different from zero in each case.

It is tempting to stop here and simply conclude that peace scholars have greater freedom of choice in studying the sizes of inter-state wars than historically presumed. While this is true, it would be a mistake to think this freedom of choice is free of consequence. In this section, two of these consequences are demonstrated.

First, different measurement (and by extension, modeling) choices yield conflicting conclusions about the so-called long peace. Many scholars argue that starting after World War II the international system entered an unprecedented period of peace, at least among major powers.[5] This argument is situated within a broader set of claims known as the decline-of-war thesis, which holds that wars over time, in addition to becoming less common, are becoming less deadly.

Some scholars answer in the affirmative on this question (Cunen, Hjort, and Nygård 2020; Pinker 2011) , while others answer in the negative (Braumoeller 2019; Clauset 2017, 2018). Most point to World War II as the relevant turning point, and Cunen, Hjort, and Nygård (2020) recently identified the year 1950 as the most statistically likely. In the analysis that follows, the 1950 cutpoint is adopted so

[5] See Braumoeller (2019) and Cunen, Hjort, and Nygård (2020) for a comprehensive set of citations and summaries.

that the results are comparable with the most recent statement made about the long peace. To illustrate how choice of measure and model influences conclusions about the long peace, a bootstrapped test like that developed by Braumoeller (2019) is conducted for the power-law slopes fitted to pre- and post-1950 battle deaths. Additional bootstrapped tests are performed using the inverse Burr and log-normal models. As the results in the next section show, a long peace is statistically undetectable with the power-law and the inverse Burr regardless of how war size is measured. However, a significant difference is identified for war death intensity using the log-normal model.

The second implication that is demonstrated is the possibility for conducting regression analysis of correlates of war size. Power-law distributed data is sometimes incompatible with regression analysis (Braumoeller 2019; Taleb 2010)—after all, it is hard to identify the conditional mean of a scale-free response. However, if the data can be justifiably treated as log-normal or inverse Burr distributed, regression analysis is back on the table.[6] So, in the analysis that follows, both inverse Burr and log-normal regression models are considered, and a few different covariates are included in the analysis.

These variables were added to the conflict series using the `{peacesceincer}` R package (Miller 2022). They include a measure of the "weakest link" in terms of democratic quality, total military personnel of the countries fighting a war, global population, and the population of the countries fighting a war. Weakest link democracy is measured as the minimum democracy score from the Varieties of Democracy project among the countries fighting a war (Coppedge et al. 2020). Population and military personnel data come from the Correlates of War National Material Capabilities dataset (J. D. Singer, Bremer, and Stuckey 1972; D. J. Singer 1987). Except for democracy, each of the measures is log-transformed to normalize the data. For the log-normal model in particular, this transformation also means that the estimates for military and population size can be interpreted as elasticities since the outcome is also log-transformed prior to estimation.

The results demonstrate that war size is not entirely random. Instead, military and population size, as well as democracy are significant predictors of how deadly wars can become. While a greater

[6] Cunen, Hjort, and Nygård (2020) demonstrate an application with the inverse Burr model.

Table 7: A test of the long-peace using the classic power-law model.

Data	pre-1950	post-1950	Difference	p-value
Severity	1.489	1.629	0.140	0.494
Prevalence	1.535	1.641	0.106	0.415
Intensity	1.673	1.667	-0.006	0.494

Entries are power-law slopes. 2,000 bootstraps performed.

number of military personnel and a larger overall global population predict a greater number of battle deaths, higher quality of democracy and a larger combined population among the countries fighting a war predict fewer deaths. Models also include a post-1950 dummy variable. Contrary to the findings regarding the long peace in the first analysis below, the results show that, controlling for democracy, population, and military size, there is a statistically significant decline in the sizes of wars beginning in the mid-twentieth century.

Identifying the "Long-Peace"

First up is the question of the long peace. Tables 7, 8, and 9 report results from statistical tests comparing pre-1950 to post-1950 trends in conflict size. The estimates in each table are based on each of the alternative models of war size. Table 7 shows estimates from power-law fits for the data. Following Braumoeller (2019), a bootstrapped test is performed to assess whether the data pre- and post-1950 have statistically different power-law slopes in their extreme tail. Tests are done with battle death severity, prevalence, and intensity. Cell entries are the different power-law slopes, their difference, and the bootstrap p-value. Though Braumoeller (2019) used 1945 as the cutoff in his analysis, the 1950 cutoff used here was identified as the most optimal by Cunen, Hjort, and Nygård (2020). This different year, however, yields similar conclusions to Braumoeller's. The power-law slopes pre- and post-1950 are all less than 2, consistent with battle deaths being scale-free regardless of the time period and how conflict size is measured. Further, the power-law slopes in the different periods are not statistically different. That means we have little evidence that the most extreme wars before and after 1950 are generated by a different power-law distribution.

Table 8: A test of the long-peace using the inverse Burr model.

Data	pre-1950	post-1950	Difference	p-value
Severity	0.636	0.000102	-0.636	0.318
Prevalence	3e-10	2.48e-07	2.48e-07	0.453
Intensity	139	0.00177	-139	0.293

Entries are central tendency for inverse Burr. 2,000 bootstraps performed.

Table 8 reports results using the inverse Burr model. As with the power-law, a bootstrap test is performed. Recall that the inverse Burr has three parameters. The analysis here homes in on the scaling parameter μ which denotes the central tendency of the inverse Burr distribution. If this parameter is different between periods, this indicates a change in the expected rate of battle deaths. Cell entries in Table 8 are estimates of μ pre- and post-1950 along with their difference and the bootstrapped p-value. Like with the power-law, across alternative measures of battle deaths the inverse Burr central tendency pre- and post-1950 is not statistically different.

Finally, Table 9 shows estimates from bootstrapped tests of the difference in the log-mean of battle deaths pre- and post-1950. As with the inverse Burr, the parameter μ (log-mean) captures the central tendency of the log-normal distribution. In the cases of battle death severity and prevalence, there is no statistically significant difference in war size pre- and post-1950. However, in the case of battle death intensity, there is. This finding is worth noting because only for intensity was the log-normal model not rejected as a plausible data-generating process for war size. In addition, though Vuong's test could not reject the null that the log-normal was superior to the power-law and inverse Burr, the signs of the test nonetheless favored the log-normal. What we should make of this is discussed in the final section of the paper.

Parameterizing Correlates of War Size

The previous section considered how measurement and model choice might influence conclusions about the long peace. In this section, regression analysis using the inverse Burr and log-normal models is demonstrated. Results are shown for each of the measures of war size using the explanatory variables:

Table 9: A test of the long-peace using the log-normal model.

Data	pre-1950	post-1950	Difference	p-value
Severity	8.365	3.258	-5.107	0.257
Prevalence	0.525	-1.623	-2.148	0.386
Intensity	5.522	3.867	-1.654	0.005

Entries are central tendency for log-normal. 2,000 bootstraps performed.

global population and pooled population of the countries fighting a war (log), military size (log), and weakest link democracy. Dummy variables for post-1950 are also included to assess how controlling for covariates might change conclusions about the long peace. For completeness, these regression models were fit for each of the measures of war size despite the fact that the inverse Burr and log-normal models could be rejected for battle death severity and the log-normal also could be rejected for battle death prevalence.[7]

Following Cunen, Hjort, and Nygård (2020) who describe how the inverse Burr model can be parameterized to incorporate covariates in its estimation, the inverse Burr regression applied here operates by modifying the scaling parameter (central tendency) μ such that

$$\mu_i = \exp(X_i^\top \beta_k) \tag{6}$$

In words, μ_i replaces the constant μ and is operationalized as the exponent of the linear combination of a vector of covariates. The exponent is used to ensure $\mu_i > 0$. To estimate the modified inverse Burr model, a standard numerical optimizer is used and bootstrapping is done for statistical inference. The R code necessary to implement the regression is available in the Code Appendix.

For the log-normal model, each of the measures of war size was log-transformed and then regressed on the set of covariates. Model parameters were estimated using OLS with robust (HC1) standard errors.

The results from the analysis are summarized in Figure 7, which is a coefficient plot of the point

[7] However, because of the identities of the model parameters, with the log of the populations of the countries fighting a war as a right-hand-side variable, each of the log-normal models are equivalent to the model of battle death intensity.

estimates and their 95% confidence intervals for each of the covariates included in the analysis. Inverse Burr estimates are in blue and log-normal (OLS) estimates are shown in red. The gray column to the left in each panel of the figure shows the value of the coefficient with stars denoting the level of statistical significance. Going from the left panel of the figure to the right, results are shown with severity, prevalence, and intensity as the outcome, respectively.

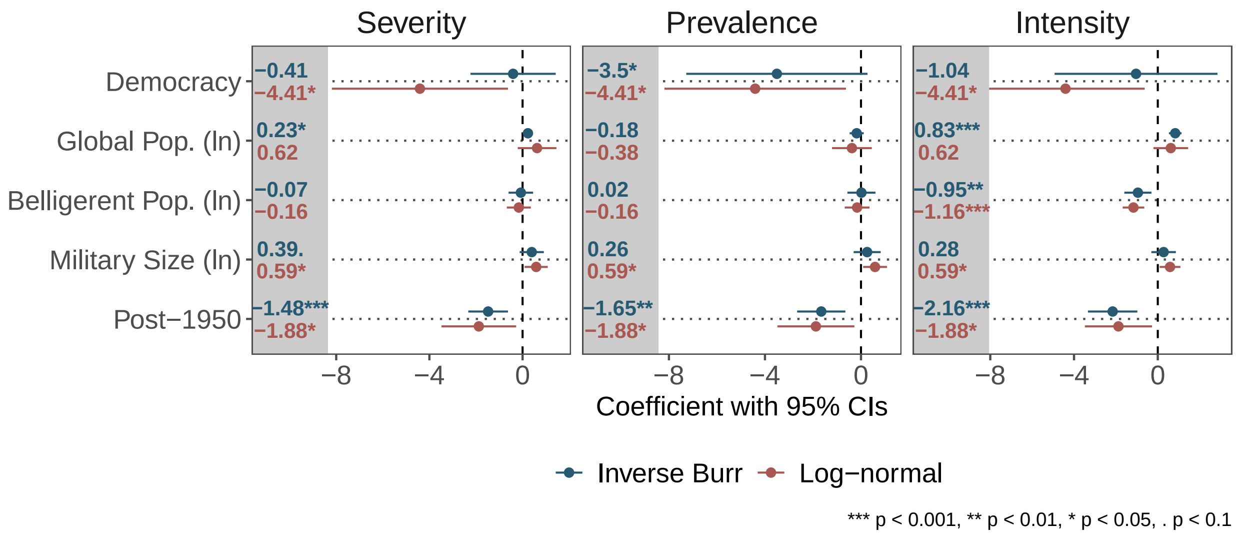

Figure 7: Estimates for parameterized inverse Burr and log-normal models.

There are few differences in the direction and significance of covariates across models and outcomes; however, we see the greatest heterogeneity across outcomes in the inverse Burr models. First, democracy only is a significant predictor of battle deaths in the case of global prevalence. The estimate shows that the better the quality of democracy in the weakest democracy among those fighting a war, the lower the central tendency of battle deaths per global population. With respect to severity and intensity, no such relationship is statistically detectable.

With respect to global population, the inverse Burr coefficient is statistically significant and positive in the case of war death severity and intensity; but not prevalence. For the first two outcomes, the estimate is positive, suggesting that as the global population has increased, the central tendency of battle deaths has increased.

With respect to the populations of the belligerent countries (those fighting a war), the inverse Burr coefficient is only statistically significant in the case of battle death intensity. The estimate is negative, indicating that the larger the overall population pool of the countries fighting a war, the smaller the number of battle deaths per capita.

Turning to military size, across all inverse Burr models the coefficient on military personnel is positive but statistically insignificant; though, for battle death severity the estimate approaches significance at the $p < 0.1$ level. This is surprising, since one would naturally expect wars fought between larger fighting forces to have more overall deadly potential.

Finally, the post-1950 indicator is statistically significant across all inverse Burr models. This runs contrary to the results shown in the previous section. This suggests that once additional covariates are controlled for in the analysis, the inverse Burr model is consistent with the long peace after all.

Next up are the estimates from the log-normal models. Unlike with the inverse Burr estimates, these are less variable across outcomes. The reason for this has to do with the simple fact that in log-log models (where both the outcome and explanatory variables are log-transformed), the identities of the model parameters are identical regardless of whether a predictor is also a denominator for the outcome. Consider that $\log(y/x) = \alpha + \beta \log(x)$ can be rearranged such that $\log(y) = \alpha + (\beta + 1)\log(x)$.[8] In practice, that means that regressing the log of the ratio of y to x on the log of x is functionally equivalent to just regressing the log of y on the log of x (the only difference will be whether the slope parameter is β or β + 1). The upshot for peace scholars studying variation in the sizes of wars is that, as long as the log of pooled belligerent population is controlled for in a regression analysis, any of the measures of war size can be justifiably studied using a log-log specified linear regression model. This is because such a regression is functionally equivalent to modeling battle death intensity—an outcome for which the log-normal model could not be rejected and for which the sign of model comparison tests suggests is a better fit for the data (albeit, the estimate was not statistically significant).

Turning the the regression estimates, in all log-normal models democracy is a significant predictor of conflict size. The better the quality of democracy in the weakest democracy fighting a war, the fewer expected battle deaths we can expect to see.

[8]Since $\log(y/x) \equiv \log(y) - \log(x)$ we can rearrange so that $\log(y) = \alpha + \beta \log(x) + \log(x)$, which reduces to $\log(y) = \alpha + (\beta + 1)\log(x)$.

For global population, the results differ by outcomes. When the outcome is conflict severity, global population is a positive predictor of conflict size. Since the results can be interpreted as elasticities, we can directly interpret the coefficient as the percentage difference in severity per a percent increase in global population. A one percent increase in global population predicts a 0.23% increase in conflict severity (total battle deaths). When modeling battle death prevalence the coefficient switches sign to negative and is statistically insignificant. However, when modeling battle death intensity, the estimate is again positive and statistically significant. Specifically, a one percent increase in global population predicts a 0.83% increase in conflict intensity (battle deaths per pooled belligerent population).

Turning from global population to just the population of the countries fighting a war (belligerent population), only for conflict intensity do we see a statistically significant relationship. In particular, a one percent increase in total belligerent population predicts a 1.16% decline in conflict intensity.

While having a larger populations predicts fewer deaths per capita, having larger fighting forces predicts more fatalities. Across models, a one percent increase in the total number of military personnel among the countries fighting a war predicts a 0.59% increase in expected battle deaths, whether the outcome is measured in terms of severity, prevalence, or intensity.

Finally, the estimate for the post-1950 dummy is statistically significant and negative. By transforming the coefficient using $100 \times [\exp(\beta) - 1]$, we can directly calculate the percentage change in conflict size post-1950 relative to pre-1950. The difference is quite substantial. The average conflict size after 1950 is nearly 85% lower than pre-1950 levels.

Taken together, these findings suggest that war size is not entirely random. Certain factors, like the quality of democracy, population size, military size, and period can be used to predict how deadly a war between or among a set of countries can become. This suggests several possibilities for conflict scholars and researchers. Not only might it be possible to test a number of hypotheses about conflict size with regression analysis, it also may be feasible to devise forecasts of conflict size based on historical data.

As a simple example, consider an out of sample prediction using the current conflict between Russia and Ukraine. If we feed the the log-normal model of conflict severity data for the Russia-Ukraine war, the model predicts 189,134 total battle deaths by the war's end with lower and upper 95% prediction interval bounds of 2,405 and 14,877,644, respectively.[9] As of April, the death toll for both sides neared 190,000—a total very close to the model's prediction. However, the war is still raging as of this writing, and given the bounds of the prediction intervals a death total of greater than 1 million by war's end would not be that surprising.

This example highlights an important point: the prediction intervals for the Russia-Ukraine war are quite wide. War is a complex phenomenon, so predicting its final death toll is hard to do with a high degree of precision. The outcome of any one battle is subject to substantial randomness. Compounded across multiple battles, the margin of error only increases. So, while the results from this analysis suggest that variation in war size is explicable in terms of the factors considered here, there's still plenty of variation left unexplained.

Recommendations

Taken together, the findings of this study show that conflict scholars have more methodological freedom of choice when analyzing the sizes of conflicts than previously thought. However, this study also shows that this freedom of choice is not free of consequence. To the contrary, how war size is measured determines the appropriateness of alternative models. These alternative models in turn determine the statistical precision with which hypotheses can be tested and whether regression analysis is justifiable. For these reasons, future studies on the sizes of international conflicts should be conducted with a fair dose of caution and transparency. Conflict scholars can exercise this caution and transparency by taking two simple actions.

First, use the best-practices outlined by Clauset, Shalizi, and Newman (2009) for studying thick-tailed distributions. Only by performing model validation and comparison tests can researchers have a data-driven justification for opting for one model over another. The analysis conducted here suggests that the power-law is not the only possible data-generating process for war size. Depending on how war size is operationalized, the inverse Burr or log-normal are just as appropriate (and possibly more

[9] This value is produced using the following transformation of the model predictions (which are on the log-scale) to get the expected battle deaths on their original scale: $\exp(\hat{x} + \text{var}[\epsilon_i] \cdot 2^{-1})$.

useful). However, in future studies scholars should not treat these results as the definitive statement on model validation and choice for studying war size. Different conclusions may follow from different datasets or from studying different kinds of conflict, such as civil wars. Furthermore, there are other models not considered here that might be just as applicable for studying war size. Finally, despite efforts to minimize mistakes in data collection, aggregation, and analysis in this study, errors may still have been made. In sum, scholars should support their modeling choices using their own data-driven assessments rather than those reported here or elsewhere.

The second action scholars should take is to be transparent about how model and measurement might influence their results. While different scholars may have valid justifications for opting for one measure (severity, prevalence, or intensity) or model (power-law, inverse Burr, or log-normal) over another, this choice can influence results. Therefore, however scholars justify their measure or model of conflict size, they should report results using the alternatives as well. This added transparency will help researchers guard against claims that their choice of measure or model was *ad hoc* rather than theoretically informed or justified on the basis of model validation and selection. On this point, in the analysis of the long peace presented earlier in this paper the long peace was statistically detectable using the log-normal model of conflict intensity. In all other instances, this finding failed to materialize. Even though it is arguable that the log-normal model is the best for studying war intensity, results using other measures and models should also be reported so readers can draw their own conclusions regarding the results.

Beyond this practical advice, this study points to an enhanced role for conflict scholars in policymaking. Though future work should examine this future, the analysis presented here shows that models like the inverse Burr and log-normal are justifiable alternatives to the power-law when studying conflict sizes. These modeling choices further support regression analysis. The regression models estimated here show that factors like minimum quality of democracy among the countries fighting a war, military size, and population are statistically significant predictors of war size; albeit, with considerable variation in death totals still left unexplained. While war is a complex phenomenon that is difficult to predict, with conflict on the rise in many parts of the world, policymakers are eager for data-informed forecasts about how deadly different conflicts are likely to become. While such forecasts will not be perfect, they will provide a rigorous and replicable basis for speculating about current or future wars—this a service that quantitative conflict scholars are well-positioned to provide.

In sum, conflict scholars need not limit themselves to using the power-law to study the sizes of international conflicts. By using the methods outlined in here, researchers can find data-driven justification for alternative models that also grant greater latitude (such as the ability to conduct regression analysis) and statistical precision. While this new-found freedom comes with consequences that should lead scholars to proceed with caution and to exercise transparency, the upshot of this freedom is an enhanced capacity to test a richer set of hypotheses helping to better explain why some international wars come to be deadlier than others.

Code Appendix

```
# ==============================
# Setup, packages, and helpers
# ==============================

knitr::opts_chunk$set(cache=TRUE, echo=FALSE,
                      message=FALSE, warning=FALSE)

## Packages
library(tidyverse)
library(estimatr)
library(texreg)
library(kableExtra)
library(lmtest)
library(sandwich)
library(poweRlaw)
library(actuar)
library(patchwork)
library(coolorrr)
theme_set(theme_test())
set_palette()

## Read in methods for inverse Burr to use with
## {poweRlaw} package:
source(
  here::here(
    "04_report",
    "00_coninvbur.R"
  )
)

# ====================================================
# Read in the data
# ====================================================
dt <- read_csv(
  here::here("01_data", "war-year.csv")
)

# ==================================
# Analysis for section:
# "The Power Law and Alternatives
# ==================================
```

```r
## In-text percentages for Fig. 1
num <- dt$batdeath[dt$batdeath <= quantile(dt$batdeath, 0.8)]
den <- dt$batdeath
pct <- round(100 * sum(num) / sum(den), 2)

## Fig. 1 summarizing the density of total battle deaths
ggplot(dt) +
  aes(x = batdeath) +
  geom_density(fill = "gray") +
  scale_x_log10(
    labels = scales::comma
  ) +
  geom_vline(
    aes(
      xintercept = quantile(batdeath, 0.8)
    )
  ) +
  annotate(
    "text",
    x = quantile(dt$batdeath, 0.8),
    y = 0.35,
    label = "80th",
    hjust = 1
  ) +
  labs(
    x = "Battle Deaths (log-10)",
    y = "Density",
    fill = NULL
  ) +
  ggpal(aes = "fill")

## Fig. 2: example plots of inverse CDFs with power-law
x <- 1:100
y <- (x / max(x))^-4
p1 <- ggplot() +
  aes(x, y) +
  geom_line(size = 1) +
  labs(
    x = "x",
    y = "Pr(X > x)",
    title = "Unadjusted Scale"
  )
p2 <- p1 +
  scale_x_log10() +
```

```r
  scale_y_log10() +
  labs(
    title = "log-log Scale"
  )
p1 + p2 &
  theme(
    axis.text = element_blank()
  )

## Fig. 3: the empirical inverse CDF shown for inensity of battle deaths
f <- function(x) rank(-x) / max(rank(-x))
ggplot(dt) +
  aes(x = batdeathpc * 1000000, y =f(batdeathpc)) +
  geom_point(color = "gray") +
  scale_x_log10(
    labels = scales::comma
  ) +
  scale_y_log10() +
  labs(
    x = "Battle Deaths per Million",
    y = "Pr(X > x)"
  )

## Fig. 4: Example inverse Burr inverse CDF
x <- 1:100
y <- (1 / (1 + exp(x)))^2
p1 <- ggplot() +
  aes(x, y) +
  geom_line(size = 1) +
  labs(
    x = "x",
    y = "Pr(X > x)",
    title = "Unadjusted Scale"
  )
p2 <- p1 +
  scale_x_log10() +
  scale_y_log10() +
  labs(
    title = "log-log Scale"
  )
p1 + p2 &
  theme(
    axis.text = element_blank()
  )
```

```r
## Fig. 5: Example log-normal inverse CDF
x <- 1:100
y <- 1 - pnorm(x, mean = 0, sd = 20)
p1 <- ggplot() +
  aes(x, y) +
  geom_line(size = 1) +
  labs(
    x = "x",
    y = "Pr(X > x)",
    title = "Unadjusted Scale"
  )
p2 <- p1 +
  scale_x_log10() +
  scale_y_log10() +
  labs(
    title = "log-log Scale"
  )
p1 + p2 &
  theme(
    axis.text = element_blank()
  )

# =================================
# Emperical analysis for section:
# "Analysis: Model Fitting"
# =================================

# the data
x <- dt$batdeath
y <- dt$batdeath / dt$wpop * 1000000
z <- dt$batdeathpc * 1000000

# fits for total war size
x1 <- conpl$new(x)
x1$setXmin(estimate_xmin(x1, xmax = 1e09))
x2 <- coninvburr$new(x)
x2$setPars(estimate_pars(x2))
x3 <- conlnorm$new(x)
x3$setPars(estimate_pars(x3))

# fits for "all cause mortality"
y1 <- conpl$new(y)
y1$setXmin(estimate_xmin(y1, xmax = 1e09))
y2 <- coninvburr$new(y)
```

```r
y2$setPars(estimate_pars(y2))
y3 <- conlnorm$new(y)
y3$setPars(estimate_pars(y3))

# fits for risk of war
z1 <- conpl$new(z)
z1$setXmin(estimate_xmin(z1, xmax = 1e09))
z2 <- coninvburr$new(z)
z2$setPars(estimate_pars(z2))
z3 <- conlnorm$new(z)
z3$setPars(estimate_pars(z3))

# Fig. 7: plot the results
par(mfcol = c(1, 3))
plot(x1, pch = 19, col = "gray",
     xlab = "War Size", ylab = "Pr(X > x)",
     main = "Severity")
lines(x3, col = "red3")
lines(x2, col = "steelblue")
lines(x1, col = "black")
plot(y1, pch = 19, col = "gray",
     xlab = "War Size", ylab = "Pr(X > x)",
     main = "Prevalence")
lines(y3, col = "red3")
lines(y2, col = "steelblue")
lines(y1, col = "black")
plot(z1, pch = 19, col = "gray",
     xlab = "War Size", ylab = "Pr(X > x)",
     main = "Intensity")
lines(z3, col = "red3")
lines(z2, col = "steelblue")
lines(z1, col = "black")
legend(
  "bottomleft",
  lty = c(1, 1, 1),
  col = c("black", "steelblue", "red3"),
  legend = c("Power-law", "Inverse Burr", "Log-normal"),
  bty = "n"
)

# ================================
# Emperical analysis for section:
# "Analysis: Goodness of Fit"
# ================================
```

```r
# GOF tests for battle death *severity* using 2,000 bootstraps
set.seed(1)
gof1 <- my_bootstrap_p(
  x1,
  threads = 4, no_of_sims = 2000,
  xmins = rep(x1$xmin, 2),
  seed = 1
)
gof2 <- my_bootstrap_p(
  x2, no_of_sims = 2000,
  seed = 1
)
gof3 <- my_bootstrap_p(
  x3, no_of_sims = 2000,
  threads = 4,
  xmins = rep(min(x), 2),
  seed = 1
)

tibble( # report in a table (Table 1)
  Model = c("Power-law", "Inverse Burr", "Log-normal"),
  GOF = c(gof1$gof, gof2$gof, gof3$gof),
  "p-value" = c(gof1$p, gof2$p, gof3$p)
) |>
  kbl(
    caption = "GOF Tests for Severity",
    booktabs = T,
    linesep = "",
    digits = 3
  ) |>
  add_footnote(
    "Based on 2,000 bootstraps.",
    notation = "none"
  )


# GOF tests for battle death *prevalence* with 2,000 bootstraps
set.seed(1)
gof1 <- my_bootstrap_p(
  y1,
  threads = 4, no_of_sims = 2000,
  xmins = rep(y1$xmin),
  seed = 1
)
```

```r
gof2 <- my_bootstrap_p(
  y2, no_of_sims = 2000,
  seed = 1
)
gof3 <- my_bootstrap_p(
  y3,
  threads = 4,
  xmins = rep(min(y), 2),
  no_of_sims = 2000,
  seed = 1
)

tibble( # report in a table (Table 2)
  Model = c("Power-law", "Inverse Burr", "Log-normal"),
  GOF = c(gof1$gof, gof2$gof, gof3$gof),
  "p-value" = c(gof1$p, gof2$p, gof3$p)
) |>
  kbl(
    caption = "GOF Tests for Prevalence",
    booktabs = T,
    linesep = "",
    digits = 3
  ) |>
  add_footnote(
    "Based on 2,000 bootstraps.",
    notation = "none"
  )


# GOF tests for battle death *intensity* with 2,000 bootstraps
set.seed(1)
gof1 <- my_bootstrap_p(
  z1,
  threads = 4,
  no_of_sims = 2000,
  xmins = rep(z1$xmin, 2),
  seed = 1
)
gof2 <- my_bootstrap_p(
  z2, no_of_sims = 2000,
  seed = 1
)
gof3 <- my_bootstrap_p(
  z3,
```

```
  threads = 4,
  no_of_sims = 2000,
  xmins = rep(min(z), 2),
  seed = 1
)

tibble( # report in table (Table 3)
  Model = c("Power-law", "Inverse Burr", "Log-normal"),
  GOF = c(gof1$gof, gof2$gof, gof3$gof),
  "p-value" = c(gof1$p, gof2$p, gof3$p)
) |>
  kbl(
    caption = "GOF Tests for Intensity",
    booktabs = T,
    linesep = "",
    digits = 3
  ) |>
  add_footnote(
    "Based on 2,000 bootstraps.",
    notation = "none"
  )


# =================================
# Emperical analysis for section:
# "Analysis: Model Comparisons"
# =================================

# A function to compare model likelihoods
# only for all X >= xmin they have in common.
simp_compare <- function(d1, d2) {
  if(d1$xmin == d2$xmin) {
    cp <- compare_distributions(d1, d2)
  } else {
    d1cpy <- d1$copy()
    d2cpy <- d2$copy()
    d1cpy$setXmin(max(d1$xmin, d2$xmin))
    d2cpy$setXmin(max(d1$xmin, d2$xmin))
    cp <- compare_distributions(d1cpy, d2cpy)
  }
  tibble(
    Models = paste0(
      class(d1)[1], " vs. ",
      class(d2)[1]
```

```
    ),
    Estimate = cp$test_statistic,
    "p-value" = cp$p_two_sided
  )
}

# perform model comparisons for battle death *severity* and
# report in a table (Table 4)
bind_rows(
  simp_compare(x1, x2),
  simp_compare(x1, x3),
  simp_compare(x2, x3)
) |>
  mutate(
    Models = c(
      "Power-law vs. Inverse Burr",
      "Power-law vs. Log-normal",
      "Inverse Burr vs. Log-normal"
    )
  ) |>
  kbl(
    caption = "Vuong's test for best fitting model for battle death severity.",
    booktabs = T,
    digits = 3,
    linesep = ""
  ) |>
  add_footnote(
    "Only non-truncated data points used for comparisons with the power-law. Full data u
    notation = "none"
  )

# perform model comparisons for battle death *prevalence* and
# report in a table (Table 5)
bind_rows(
  simp_compare(y1, y2),
  simp_compare(y1, y3),
  simp_compare(y2, y3)
) |>
  mutate(
    Models = c(
      "Power-law vs. Inverse Burr",
      "Power-law vs. Log-normal",
      "Inverse Burr vs. Log-normal"
    )
```

```r
  ) |>
  kbl(
    caption = "Vuong's test for best fitting model for battle death prevalence.",
    booktabs = T,
    digits = 3,
    linesep = ""
  ) |>
  add_footnote(
    "Only non-truncated data points used for comparisons with the power-law. Full data u
    notation = "none"
  )

# perform model comparisons for battle death *intensity* and
# report in a table (Table 6)
bind_rows(
  simp_compare(z1, z2),
  simp_compare(z1, z3),
  simp_compare(z2, z3)
) |>
  mutate(
    Models = c(
      "Power-law vs. Inverse Burr",
      "Power-law vs. Log-normal",
      "Inverse Burr vs. Log-normal"
    )
  ) |>
  kbl(
    caption = "Vuong's test for best fitting model for battle deaths intensity.",
    booktabs = T,
    digits = 3,
    linesep = ""
  ) |>
  add_footnote(
    "Only non-truncated data points used for comparisons with the power-law. Full data u
    notation = "none"
  )

# ==============================================
# Empirical analysis for section:
# "Implications: Identifying the 'Long-Peace'"
# ==============================================

# set up for empirical test of the long peace
library(furrr) # use parallel computing
```

```r
plan(multicore, sessions = 7) # 7 cores on my Intel i7 machine

# a wrapper for fitting and boostrapping the power-law
pl_fit <- function(dat, its = 2000) {
  # fit the model
  m <- conpl$new(dat)
  m$setXmin(estimate_xmin(m, xmax = 1e09))
  # perform bootstrap
  tibble(
    it = 1:its,
    bm = future_map(
      it, ~ {
        sdat <- sample(dat, length(dat), T)
        bm <- conpl$new(sdat)
        bm$setXmin(estimate_xmin(bm, xmax = 1e09))
        tibble(par = bm$pars)
      },
      .options = furrr_options(seed = T)
    )
  ) -> boot_out
  # return fit and bootstrap
  list(
    pars = m$pars,
    boot_pars = boot_out |>
      unnest(bm)
  )
}

# a wrapper for fitting and bootstrapping the inverse Burr
ib_fit <- function(dat, its = 2000) {
  m <- coninvburr$new(dat)
  m$setPars(estimate_pars(m))
  tibble(
    it = 1:its,
    bm = future_map(
      it, ~ {
        sdat <- sample(dat, length(dat), T)
        bm <- coninvburr$new(sdat)
        bm$setPars(estimate_pars(bm))
        tibble(
          par1 = bm$pars[1],
          par2 = bm$pars[2],
          par3 = bm$pars[3]
        )
```

```
      },
      .options = furrr_options(seed = T)
    )
  ) -> boot_out
  list(
    pars = m$pars,
    boot_pars = boot_out |>
      unnest(bm)
  )
}

# a wrapper for fitting and bootstrapping the log-normal
ln_fit <- function(dat, its = 2000) {
  m <- conlnorm$new(dat)
  m$setPars(estimate_pars(m))
  tibble(
    it = 1:its,
    bm = future_map(
      it, ~ {
        sdat <- sample(dat, length(dat), T)
        bm <- conlnorm$new(sdat)
        bm$setPars(estimate_pars(bm))
        tibble(
          par1 = bm$pars[1],
          par2 = bm$pars[2]
        )
      },
      .options = furrr_options(seed = T)
    )
  ) -> boot_out
  list(
    pars = m$pars,
    boot_pars = boot_out |>
      unnest(bm)
  )
}

# a quick function to compute p-values from bootstraps
get_p <- function(x, y) {
  2 * min(
    mean(x > y),
    mean(x < y)
  )
}
```

```r
# divide the data by pre- and post-1950
  # battle death *severity*
xpre <- x[dt$year <= 1950]
xpos <- x[dt$year > 1950]

  # battle death *prevalence*
ypre <- y[dt$year <= 1950]
ypos <- y[dt$year > 1950]

  # battle death *intensity*
zpre <- z[dt$year <= 1950]
zpos <- z[dt$year > 1950]

# power-law fits
set.seed(1)
x1pre <- pl_fit(xpre)
x1pos <- pl_fit(xpos)
y1pre <- pl_fit(ypre)
y1pos <- pl_fit(ypos)
z1pre <- pl_fit(zpre)
z1pos <- pl_fit(zpos)

# summarize results
pl_tests <- tibble(
  Data = c("Severity", "Prevalence", "Intensity"),
  "pre-1950" = c(x1pre$pars, y1pre$pars, z1pre$pars),
  "post-1950" = c(x1pos$pars, y1pos$pars, z1pos$pars),
  Difference = `post-1950` - `pre-1950`,
  "p-value" = c(
    get_p(x1pre$boot_pars$par, x1pos$boot_pars$par),
    get_p(y1pre$boot_pars$par, y1pos$boot_pars$par),
    get_p(z1pre$boot_pars$par, z1pos$boot_pars$par)
  )
)

# report in a table (Table 7)
kbl(
  pl_tests,
  caption = "A test of the long-peace using the classic power-law model.",
  digits = 3,
  booktabs = T,
  linesep = ""
) |>
  add_footnote(
```

```r
    "Entries are power-law slopes. 2,000 bootstraps performed.",
    notation = "none"
  )

# inverse Burr fits
x2pre <- ib_fit(xpre)
x2pos <- ib_fit(xpos)
y2pre <- ib_fit(ypre)
y2pos <- ib_fit(ypos)
z2pre <- ib_fit(zpre)
z2pos <- ib_fit(zpos)

# summarize results
ib_tests <- tibble(
  Data = c("Severity", "Prevalence", "Intensity"),
  "pre-1950" = c(x2pre$pars[3], y2pre$pars[3], z2pre$pars[3]),
  "post-1950" = c(x2pos$pars[3], y2pos$pars[3], z2pos$pars[3]),
  Difference = `post-1950` - `pre-1950`,
  "p-value" = c(
    get_p(x2pre$boot_pars$par3, x2pos$boot_pars$par3),
    get_p(y2pre$boot_pars$par3, y2pos$boot_pars$par3),
    get_p(z2pre$boot_pars$par3, z2pos$boot_pars$par3)
  )
)

# report in a table (Table 8)
kbl(
  ib_tests |> mutate(
    across(
      2:4, ~ signif(.x, 3) |> as.character()
    )
  ),
  caption = "A test of the long-peace using the inverse Burr model.",
  digits = 3,
  booktabs = T,
  linesep = ""
) |>
  add_footnote(
    "Entries are central tendency for inverse Burr. 2,000 bootstraps performed.",
    notation = "none"
  )

# log-normal fits
x3pre <- ln_fit(xpre)
```

```r
x3pos <- ln_fit(xpos)
y3pre <- ln_fit(ypre)
y3pos <- ln_fit(ypos)
z3pre <- ln_fit(zpre)
z3pos <- ln_fit(zpos)

# summarize results
ln_tests <- tibble(
  Data = c("Severity", "Prevalence", "Intensity"),
  "pre-1950" = c(x3pre$pars[1], y3pre$pars[1], z3pre$pars[1]),
  "post-1950" = c(x3pos$pars[1], y3pos$pars[1], z3pos$pars[1]),
  Difference = `post-1950` - `pre-1950`,
  "p-value" = c(
    get_p(x3pre$boot_pars$par1, x3pos$boot_pars$par1),
    get_p(y3pre$boot_pars$par1, y3pos$boot_pars$par1),
    get_p(z3pre$boot_pars$par1, z3pos$boot_pars$par1)
  )
)

# report in a table (Table 9)
kbl(
  ln_tests,
  caption = "A test of the long-peace using the log-normal model.",
  digits = 3,
  booktabs = T,
  linesep = ""
) |>
  add_footnote(
    "Entries are central tendency for log-normal. 2,000 bootstraps performed.",
    notation = "none"
  )

# ========================================================
# Empirical analysis for section:
# "Implications: Parameterizing Correlates of War Size"
# ========================================================

# Make a function that will estimate an inverse Burr regression
inbur_reg <- function(formula, data = NULL, its = 2000) {

  ## The Data
  y  <- model.frame(formula, data)[, 1]
  x  <- model.matrix(formula, data)
```

```r
## The likelihood
inbur_lik <- function(x, y, pars) {
  b <- matrix(
    data = pars[1:ncol(x)],
    nrow = ncol(x),
    ncol = 1
  )
  mu <- exp(x %*% b)
  alpha <- exp(pars[ncol(x) + 1])
  theta <- exp(pars[ncol(x) + 2])
  sum(
    - log(
      dinvburr(
        y,
        shape1 = alpha,
        shape2 = theta,
        scale = mu
      )
    )
  )
}

## Estimation
optim(
  par = rep(0, len = ncol(x) + 2),
  fn = inbur_lik,
  x = x,
  y = y,
  hessian = F
) -> opt_out

## Bootstrapping
tibble(
  its = 1:its,
  bout = future_map(
    its,
    ~ {
      bkeep <- sample(1:nrow(x), nrow(x), T)
      optim(
        par = rep(0, len = ncol(x) + 2),
        fn = inbur_lik,
        x = x[bkeep, ],
        y = y[bkeep],
        hessian = F
```

```r
        ) -> opt_out
        tibble(
          pars = 1:length(opt_out$par),
          vals = opt_out$par
        )
      },
      .options = furrr_options(seed = T)
    )
  ) |>
    unnest(cols = bout) |>
    group_by(pars) |>
    summarize(
      std.error = sd(vals, na.rm=T)
    ) -> boot_se

  list(
    out = tibble(
      term = c(colnames(x),
               "log(alpha)","log(theta)"),
      estimate = opt_out$par,
      std.error = boot_se$std.error,
      statistic = estimate / std.error,
      p.value = 1 - pnorm(
        abs(statistic)
      ) |> round(3)
    ),
    logLik = -opt_out$value,
    dat = model.frame(formula, data)
  )
}

# now estimate models with bootstrapped SEs:
  # the right-hand side, which includes:
  # - log of world population
  # - log of combined population of countries fighting a war
  # - log of military personnel
  # - worst V-Dem score among fighting countries
  # - post 1950 indicator
rhs <- ~ log(wpop) + log(tpop) +
  log(milper) + min_polyarchy + I(year > 1950)

  # estimate inverse Burr models for each outcome
set.seed(1)
ib_fit1 <- inbur_reg(
```

```r
  update(rhs, batdeath ~ .), data = dt
)$out
ib_fit2 <- inbur_reg(
  update(rhs, 1e06 * batdeath / wpop ~ .), data = dt
)$out
ib_fit3 <- inbur_reg(
  update(rhs, 1e06 * batdeath / tpop ~ .), data = dt
)$out

  # estimate the log-linear models via OLS with robust SEs
ln_fit1 <- lm_robust(
  update(rhs, log(batdeath) ~ .), data = dt
) |> tidy()
ln_fit2 <- lm_robust(
  update(rhs, log(1e06 * batdeath / wpop) ~ .), data = dt
) |> tidy()
ln_fit3 <- lm_robust(
  update(rhs, log(1e06 * batdeath / tpop) ~ .), data = dt
) |> tidy()

  # report the results as a coefficient plot
set_palette(
  binary = qual[c(1, 4)],
  from_coolors = F
)
bind_rows(
  ib_fit1 |> mutate(model = "Inverse Burr",
                    outcome = "Severity"),
  ib_fit2 |> mutate(model = "Inverse Burr",
                    outcome = "Prevalence"),
  ib_fit3 |> mutate(model = "Inverse Burr",
                    outcome = "Intensity"),
  ln_fit1 |> mutate(model = "Log-normal",
                    outcome = "Severity"),
  ln_fit2 |> mutate(model = "Log-normal",
                    outcome = "Prevalence"),
  ln_fit3 |> mutate(model = "Log-normal",
                    outcome = "Intensity")
) |> filter(
  !(term %in% c("log(alpha)", "log(theta)", "(Intercept)"))
) |>
  mutate(
    outcome = factor(
      outcome, levels = c("Severity", "Prevalence", "Intensity")
```

```r
    )
  ) |>
  select(term:p.value, model:outcome) |>
  ggplot() +
  aes(
    x = estimate,
    y = term,
    xmin = estimate - 1.96 * std.error,
    xmax = estimate + 1.96 * std.error,
    color = model
  ) +
  geom_vline(
    xintercept = 0,
    linetype = 2
  ) +
  geom_pointrange(
    position = ggstance::position_dodgev(-.5),
    size = .3
  ) +
  geom_vline(
    xintercept = -11,
    color = "gray80",
    linewidth = 30
  ) +
  geom_text(
    aes(x = -11,
        label = paste0(round(estimate, 2),
                       gtools::stars.pval(p.value))),
    position = ggstance::position_dodgev(-.75),
    show.legend = F,
    fontface = "bold",
    hjust = 0.2
  ) +
  facet_wrap(~ outcome, scales = "free_x") +
  ggpal(type = "binary") +
  scale_y_discrete(
    labels = c("Post-1950", "Military Size (ln)",
               "Belligerent Pop. (ln)", "Global Pop. (ln)",
               "Democracy")
  ) +
  labs(
    x = "Coefficient with 95% CIs",
    y = NULL,
    color = NULL,
```

```r
    caption = "*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1"
  ) +
  ggthemes::theme_few() +
  theme(
    legend.position = "bottom",
    strip.text = element_text(
      size = 16
    ),
    axis.text = element_text(
      size = 14
    ),
    axis.title = element_text(
      size = 14
    ),
    legend.text = element_text(
      size = 14
    ),
    panel.grid.major.y = element_line(
      linetype = 3,
      color = "gray30"
    )
  )

# ===============================================
# A numerical example using Russia and Ukraine
# ===============================================

ln_fit3 <- lm_robust(
  update(rhs, log(batdeath) ~ .), data = dt
)
newdt <- tibble(
  min_polyarchy = 0.21,
  wpop = 7.888 * 1e09,
  tpop = (143.4 + 43.79) * 1e06,
  milper = 1.454 * 1e06,
  year = 2021
)
predict(
  ln_fit3,
  newdata = newdt,
  se.fit = T,
  interval = "prediction"
)$fit |>
  apply(1, function(x) {
```

```
    scales::comma(exp(x) + sqrt(ln_fit3$res_var) / 2)
  }) -> preds
```

References

Braumoeller, Bear F. 2019. *Only the Dead: The Persistence of War in the Modern Age*. New York: Oxford University Press.

Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *American Political Science Review* 97 (1): 135–59.

Cederman, Lars-Erik, T. Camber Warren, and Didier Sornette. 2011. "Testing Clausewitz: Nationalism, Mass Mobilization, and the Severity of War." *International Organization* 65 (4): 605–38.

Cirillo, Pasquale, and Nassim Nicholas Taleb. 2016. "On the Statistical Properties and Tail Risk of Violent Conflicts." *Physica A: Statistical Mechanics and Its Applications* 452: 29–45.

Clauset, Aaron. 2017. "The Enduring Threat of a Large Interstate War." Technical report. One Earth Foundation.

———. 2018. "Trends and Fluctuations in the Severity of Interstate Wars." *Science Advances* 4 (2): eaao3580.

Clauset, Aaron, Cosma Rohilla Shalizi, and M.E.J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, et al. 2020. "V-Dem Codebook V10."

Cunen, Céline, Nils Lid Hjort, and Håvard Mokleiv Nygård. 2020. "Statistical Sightings of Better Angels: Analysing the Distribution of Battle-Deaths in Interstate Conflict over Time." *Journal of Peace Research* 57 (2): 221–34.

Fazio, Giorgio, and Marco Modica. 2015. "Pareto or Log-Normal? Best Fit and Truncation in the Distribution of All Cities." *Journal of Regional Science* 55 (5): 736–56.

González-Val, Rafael. 2016. "War Size Distribution: Empirical Regularities Behind Conflicts." *Defence and Peace Economics* 27 (6): 838–53.

Miller, Steven V. 2022. "peacesciencer: An r Package for Quantitative Peace Science Research." *Conflict Management and Peace Science*. http://svmiller.com/peacesciencer/.

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking Adult.

Reiter, Dan, Allan C Stam, and Michael C Horowitz. 2016. "A Deeper Look at Interstate War Data: Interstate War Data Version 1.1." *Research & Politics* 3 (4): 2053168016683840.

Richardson, Lewis F. 1948. "Variation of the Frequency of Fatal Quarels with Magnitude." *American Statistical Association* 43: 523–46.

———. 1960. *Statistics of Deadly Quarels*. Chicago: Quadrangle Books.

Sarkees, Meredith Reid, and Frank Wayman. 2010. *Resort to War: 1816 - 2007*. Washington DC: CQ Press.

Singer, David J. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985." *International Interactions* 14 (1): 115–32.

Singer, J. David, Stuart A. Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War and Numbers*, edited by Bruce Russett. Beverly Hills, CA: Sage Publications, Inc.

Taleb, Nassim Nicholas. 2010. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Verbeeck, Cis, Emil Kraaikamp, Daniel F. Ryan, and Olena Podladchikova. 2019. "Solar Flare Distributions: Lognormal Instead of Power Law?" *The Astrophysical Journal* 884 (50): 1–16.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica: Journal of the Econometric Society*, 307–33.

Zwetsloot, Remco. 2018. "Testing Richardson's Law: A (Cautionary) Note on Power Laws in Violence Data" Working Paper.